

PDF RENDERIZATION OF Example.R OUTPUT

```
library(data.table)
library(OaxacaSurvey)
library(fastDummies)

## Thank you for using fastDummies!

## To acknowledge our work, please cite the package:

## Kaplan, J. & Schlegel, B. (2023). fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from

# Import dataset with multilevel categorical data from a s
df <- fread("tests/eff-pool-2002-2020.csv")
df <- df[sv_year == 2020]
df[, group := 0][class == "worker", group := 0][class == "capitalist", group := 1]
df[, rentsbi := 0][rents >= renthog * 0.1 & rents > 2000, rentsbi := 1]
df[homeowner == "", homeowner := "Non-Owner"]
df$class <- relevel(as.factor(df$class), ref = "self-employed")
df$bage <- relevel(as.factor(df$bage), ref = "0-34")
df$inherit <- relevel(as.factor(df$inherit), ref = "Non-inherit")
df$homeowner <- relevel(as.factor(df$homeowner), ref = "Non-Owner")
df$riquezafin <- factor(as.logical(df$riquezafin), levels = c(TRUE, FALSE), labels = c("Fin", "NonFin"))
total_variables <- c(
  "facine3", "renthog", "renthog1", "bage", "homeowner", "worker", "young", "sex", "class",
  "actreales", "riquezanet", "riquezafin", "educ", "auton", "class",
  "tipo_auton", "direc", "multipr", "useprop", "inherit"
)
selected_variables <- c(
  "renthog", "bage", "sex", "homeowner", "riquezafin"
)

#####

# select regressors
data <- df[, ..selected_variables]

# transform multi-level categories to dummies. IMPORTANT remove both first dummy and selected column to
data <- fastDummies::dummy_cols(data,
  select_columns = c("bage", "sex", "homeowner", "riquezafin"),
  remove_first_dummy = TRUE,
  remove_selected_columns = TRUE
)

# name the columns as x1...xn for oaxaca-blinder survey
colnames(data) <- paste0("x", seq_along(colnames(data)))
length_reg <- length(colnames(data))
new_formula <- paste("y ~", paste0("x", 1:length_reg, collapse = " + "))

# compose the final dataset by adding: engenuous variable, groups of split and sample weights
data <- cbind(y = df$rentsbi, group = df$group, weights = df$facine3, data)

# finally, test the model
result2 <- oaxaca_blinder_svy(
  as.formula(new_formula),
```

```

data = data.frame(data),
group = "group",
weights = "weights",
R = 10
)
print(result2)

```

```

##          unex          end          coef          inter          total  means1_y
## value -0.003979399 -0.004826978 0.15863546 -0.11553744 0.034291649 0.10422011
## CI1   -0.052619308 -0.011298278 0.01044453 -0.18491961 -0.008686063 0.06031712
## CI2    0.084295110 0.001126774 0.24617690 0.03678725 0.078743847 0.14487167
##          means2_y  means_dif
## value 0.06992846 0.034291649
## CI1 0.06477536 -0.008686063
## CI2 0.07654812 0.078743847

```